

MULTIVARIATE COMPOUNDS WITH EQUAL NUMBER OF SUMMANDS

Pavlina K. Jordanova ¹

The paper considers multivariate discrete random sums with equal number of summands. Such distributions describe the total claim amount received by a company in a fixed time point. In Queuing theory they characterize cumulative waiting times of customers up to time t . In Theory of branching processes they model the number of heirs at a fixed point in time.

Here some general properties and formulae for numerical characteristics of these distributions are derived and some particular cases are considered.

1. Introduction

Univariate random sums, are considered intensively in the beginning of the last century. Abraham Wald [17] found the relation between the numerical characteristics of these random sums, their summands and their number of summands. Nowadays his results are well known as "Wald equalities". Investigation of the particular cases starts with Poisson distributed number of summands and geometrically distributed on positive integers summands. This case is related with the names of Alfred Aeppli (1924) and George Pólya (1930). An exploration of univariate compound Poisson processes with their applications in insurance can be seen e.g. in the works of O. Lundberg [10] (1940).

¹The author is partially supported by Project RD-08-69/02.02.2016 from the Scientific Research Fund in Konstantin Preslavsky University of Shumen, Bulgaria.

2000 *Mathematics Subject Classification*: 62E15

Key words: Compound distributions; Multinomial distribution; Negative multinomial distribution

Here we consider the corresponding general multivariate distributions. Along the paper " $\stackrel{d}{=}$ " stands for the equality in distribution of two random variables (r.v.s.), " \equiv " for coincidence of two classes of probability laws, $FI X := \frac{Var X}{E X}$ is the Fisher index of dispersion, $CV(X, Y) = \sqrt{\frac{cov(X, Y)}{E X E Y}}$, and $CV X = \frac{\sqrt{Var X}}{E X}$ is the coefficient of variation of the distribution of the random variable (r.v.) X . The sign " \sim " expresses the fact, that the distribution of the corresponding r.v. belongs to a given class of probability laws.

Let N be a non-negative, integer valued r.v. with probability generating function (p.g.f.) G_N . Suppose $(Y_{1i}, Y_{2i}, \dots, Y_{ki})$, $i = 1, 2, \dots$ are independent identically distributed (i.i.d.) random vectors, defined on the same probability space (Ω, \mathcal{A}, P) . Denote their p.g.f. by $G_{Y_{1i}, Y_{2i}, \dots, Y_{ki}}$, $i = 1, 2, \dots$. The random vector $(X_{1N}, X_{2N}, \dots, X_{kN})$, which coordinates are defined by

$$(1) \quad X_{sN} = I_{N>0} \sum_{i=1}^N Y_{si}, \quad s = 1, 2, \dots, k$$

is called **compound $N(N$ -stopped sum, random sum) with equal number of Y summands**. In any particular case, the letters N and Y , will be replaced by the name of the corresponding distribution. Briefly we will denote this in the following way

$$(X_{1N}, X_{2N}, \dots, X_{kN}) \sim C N Y(\vec{a}_N; \vec{b}_Y).$$

Here the letters $C N Y(\vec{a}_N; \vec{b}_Y)$ have the following meaning: C comes from "compound", N is the abbreviation of the distribution of the number of summands, with parameters \vec{a}_N and Y is the abbreviation of the distribution of the vector of summands, with parameters \vec{b}_Y .

The cases when N is Poisson distributed and $(Y_{1i}, Y_{2i}, \dots, Y_{ki})$ are one of the

- Multinomial(Mn),
- Negative Multinomial(NMn),
- Multivariate Geometric(MGe),
- Multivariate Poisson (MPo)

distributed are investigated in 1962, by G. Smith [14]. The particular case, when the summands are MGe on positive integers and the number of summands are Poisson, is investigated in series of papers of Minkova et. al, see e.g. [11] and

[12]. In 1971, Khatri [9] mentioned that the class of these distributions is too wide and investigates some multivariate particular cases. In 1981 Charalambides [2], together with Papageorgiou [3] explored bivariate case. Charalambides [2] has called all distributions with p.g.f. $G_N(G_{Y_{11}, Y_{21}}(z_1, z_2))$ "generalized", and has mentioned that many bivariate contagious and compound distributions belong to this class. In 1996 Wang [18], obtained multivariate compound Poisson distributions as the limiting distributions of multivariate sums of independent random vectors. He called the distribution, defined in (1), "multivariate compound distribution of type I". The covariances formula for the general case of these distributions

$$(2) \quad \text{cov}(X_{iN}, X_{jN}) = EN \text{cov}(Y_{i1}, Y_{j1}) + EY_{i1}EY_{j1} \text{Var}N$$

can be found in Sundt et. al [16]. In series of papers they consider recursions formulae for these distributions.

This paper develops the results in Jordanova [8]. Their proves are presented and new explicit relations between the probability mass functions (p.m.fs.) and the mean square regressions in particular cases are obtained. After the characterisation, some simulations of realizations of random vectors with such distributions are made, and the dependence between their coordinates is visualised. These observations show that these distributions are appropriate for modelling of both: linear dependence between the coordinates and clustering in the observations. They can serve to overcome Simpson's paradox, described e.g. in Colin [4]. The main machinery of probability generating functions (p.g.fs.) used here can be seen e.g. in the book of Johnson et. al [7].

In the next section we investigate the general properties of these distributions without specifying the probability laws of the summands in (1), or those of the number of summands. Then, in Section 3, the particular cases of Mn and NMn summands are considered. The paper finishes with some conclusive remarks.

2. General properties of multivariate compound distributions

In this section we investigate the main properties of the distribution of the multivariate random sums, defined in (1), together with their numerical characteristics and conditional distributions.

Theorem 1. *Let $X_{1N}, X_{2N}, \dots, X_{kN}$ be a random vector, defined in (1). Its distribution possesses the following properties.*

1. *Characterisation of the joint probability distribution.*

If the p.g.fs. exist in (z_1, z_2, \dots, z_k) , then

$$(3) \quad G_{X_{1N}, X_{2N}, \dots, X_{kN}}(z_1, z_2, \dots, z_k) = G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(z_1, z_2, \dots, z_k)).$$

Its probability mass function (p.m.f.) satisfies the following equalities

$$(4) \quad \begin{aligned} P(X_{1N} = 0, X_{2N} = 0, \dots, X_{kN} = 0) &= G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(0, 0, \dots, 0)) = \\ &= G_N(P(Y_{11} = 0, Y_{21} = 0, \dots, Y_{k1} = 0)). \end{aligned}$$

If the following derivatives exist, then for $s = 1, 2, \dots, k$, $x_s = 0, 1, \dots$:
 $(x_1, x_2, \dots, x_k) \neq (0, 0, \dots, 0)$,

$$(5) \quad \begin{aligned} P(X_{1N} = x_1, X_{2N} = x_2, \dots, X_{kN} = x_k) &= \\ &= \frac{\partial^{x_1+x_2+\dots+x_k} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(z_1, z_2, \dots, z_k))}{x_1! x_2! \dots x_k! \partial z_1^{x_1} \partial z_2^{x_2} \dots \partial z_k^{x_k}} \Big|_{(z_1, z_2, \dots, z_k) = (0, 0, \dots, 0)} = \end{aligned}$$

$$(6) \quad \begin{aligned} &= \sum_{n=1}^{\infty} P(N = n) \sum_{i_{11}+i_{12}+\dots+i_{1n}=x_1} \sum_{i_{21}+i_{22}+\dots+i_{2n}=x_2} \dots \sum_{i_{k1}+i_{k2}+\dots+i_{kn}=x_k} \\ &\quad \prod_{s=1}^n P(Y_{1s} = i_{1s}, Y_{2s} = i_{2s}, \dots, Y_{ks} = i_{ks}). \end{aligned}$$

2. Characterisation of subsets of the coordinates.

For all $r = 1, 2, \dots, k$ and for all subset of coordinates $X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}$,

$$(7) \quad G_{X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}}(z_{i_1}, z_{i_2}, \dots, z_{i_r}) = G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(\vec{z}_{\vec{i}})),$$

where the vector $\vec{z}_{\vec{i}}$ has coordinates equal to 1 on the places different from i_1, i_2, \dots, i_r and correspondingly coordinates $z_{i_1}, z_{i_2}, \dots, z_{i_r}$ on the other places.

$$(8) \quad \begin{aligned} P(X_{i_1N} = x_{i_1}, X_{i_2N} = x_{i_2}, \dots, X_{i_rN} = x_{i_r}) &= \\ &= \frac{\partial^{x_{i_1}+x_{i_2}+\dots+x_{i_r}} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(\vec{z}_{\vec{i}}))}{x_{i_1}! x_{i_2}! \dots x_{i_r}! \partial z_{i_1}^{x_{i_1}} \partial z_{i_2}^{x_{i_2}} \dots \partial z_{i_r}^{x_{i_r}}} \Big|_{(z_{i_1}, z_{i_2}, \dots, z_{i_r}) = (0, 0, \dots, 0)}. \end{aligned}$$

3. *Marginal distributions and their dependence on N . For $i = 1, 2, \dots, k$,*

$$(9) \quad G_{X_{iN}}(z_i) = G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(1, 1, \dots, 1, z_i, 1, \dots, 1)),$$

$$P(X_{iN} = 0) = G_N(P(Y_{i1} = 0)).$$

For $x_i \neq 0$, $P(X_{iN} = x_i) =$

$$(10) \quad = \sum_{n=1}^{\infty} P(N = n) \sum_{j_1 + j_2 + \dots + j_n = x_i} P(Y_{i1} = j_1) P(Y_{i1} = j_2) \dots P(Y_{i1} = j_n).$$

If $E Y_{i1} < \infty$ and $E N < \infty$, then

$$(11) \quad E X_{iN} = E N E Y_{i1}, \quad E(X_{iN} | N = n) = n E Y_{i1}.$$

If $E(Y_{i1}^2) < \infty$ and $E(N^2) < \infty$, then

$$(12) \quad \text{Var } X_{iN} = \text{Var } N (E Y_{i1})^2 + E N \text{Var } Y_{i1}, \quad F I X_{iN} = F I N E Y_{i1} + F I Y_{i1}$$

$$(13) \quad \text{cov}(X_{iN}, N) = E Y_{i1} \text{Var } N,$$

$$\text{cor}(X_{iN}, N) =$$

$$(14) \quad = E Y_{i1} \sqrt{\frac{\text{Var } N}{\text{Var } N (E Y_{i1})^2 + E N \text{Var } Y_{i1}}} = \sqrt{\frac{F I N}{F I N + (C V Y_{i1})^2}}.$$

4. *Conditional distributions.*

$$(15) \quad P(X_{iN} = x_i | X_{jN} = x_j) = \frac{1}{x_i!} \frac{\frac{\partial^{x_i+x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(\vec{z}_{ij\bar{i}\bar{j}}))}{\partial z_i^{x_i} \partial z_j^{x_j}} \big|_{(z_i, z_j)=(0,0)}}{\frac{\partial^{x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(1, \dots, 1, z_j, 1, \dots, 1))}{\partial z_i^{x_i}} \big|_{z_j=0}},$$

where the vector $\vec{z}_{ij\bar{i}\bar{j}}$ has coordinates equal to 1 on the places different from i, j and correspondingly coordinates z_i, z_j , on the places i and j .

$$(16) \quad G_{X_{iN}}(z_i | X_{jN} = x_j) = \frac{\frac{\partial^{x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(\vec{z}_{ij\bar{i}\bar{j}}))}{\partial z_j^{x_j}}}{\frac{\partial^{x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(1, \dots, 1, z_j, 1, \dots, 1))}{\partial z_j^{x_j}}} \big|_{z_j=0},$$

$$(17) \quad E(X_{iN}|X_{jN} = x_j) = \frac{\frac{\partial}{\partial z_i} \left[\frac{\partial^{x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(\bar{z}_{ij\bar{j}}))}{\partial z_j^{x_j}} \Big|_{z_j=0} \right] \Big|_{z_i=1}}{\frac{\partial^{x_j} G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(1, \dots, 1, z_j, 1, \dots, 1))}{\partial z_j^{x_j}} \Big|_{z_j=0}},$$

5. *Dependence between the coordinates.*

$$(18) \quad \begin{aligned} \text{cor}(X_{iN}, X_{jN}) &= \\ &= \frac{EN \text{cov}(Y_{i1}, Y_{j1}) + EY_{i1}EY_{j1}Var N}{\sqrt{[Var N (EY_{i1})^2 + ENVar Y_{i1}][Var N (EY_{j1})^2 + ENVar Y_{j1}]} } = \\ &= \frac{[CV(Y_{i1}, Y_{j1})]^2 + FIN}{\sqrt{[(CV Y_{i1})^2 + FIN][(CV Y_{j1})^2 + FIN]}}. \end{aligned}$$

6. *Distributions of the sum of the coordinates.*

$$(19) \quad G_{X_{1N}+X_{2N}+\dots+X_{kN}}(z) = G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(z, z, \dots, z)),$$

$$(20) \quad P(X_{1N}+X_{2N}+\dots+X_{kN} = s) = \frac{\partial^s G_N(G_{Y_{11}, Y_{21}, \dots, Y_{k1}}(z, z, \dots, z))}{s! \partial z^s} \Big|_{z=0}.$$

7. *Marginal distributions, given the sum of the coordinates. For $m = 0, 1, \dots, s$,*

$$(21) \quad \begin{aligned} P(X_{iN} = m | X_{1N} + X_{2N} + \dots + X_{kN} = s) &= \\ &= \binom{s}{m} \frac{\frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{i-11}, Y_{i1}, Y_{i+11}, \dots, Y_{k1}}(z, \dots, z, z_i, z, \dots, z))}{\partial z^{s-m} \partial z_i^m} \Big|_{(z, z_i)=(0,0)}}{\frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{k1}}(z, \dots, z))}{\partial z^s} \Big|_{z=0}}, \\ G_{X_{iN}}(z_i | X_{1N} + X_{2N} + \dots + X_{kN} = s) &= \\ &= \binom{s}{m} \frac{\sum_{m=0}^s z_i^m \frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{i-11}, Y_{i1}, Y_{i+11}, \dots, Y_{k1}}(z, \dots, z, z_i, z, \dots, z))}{\partial z^{s-m} \partial z_i^m} \Big|_{(z, z_i)=(0,0)}}{\frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{k1}}(z, \dots, z))}{\partial z^s} \Big|_{z=0}}. \end{aligned}$$

Proof.

1. Equality (3) follows by the definition of p.g.f. and the Fubini's theorem, in particular also known as a "Double expectations formula". Double application of equality (3) for $(z_1, z_2, \dots, z_k) = (0, 0, \dots, 0)$, (see e.g. (34.2) in [7]) leads us to (4). The main properties of p.g.f., see e.g. p.327 in [6], together with (3) entails (5). The total probability formula leads us to (6).

2. The property (7) follows by (3) and the definition of p.g.f. See e.g. p. 3 in Johnson et. al [7]. The proof of the (8) is analogous to (5).
3. (9) and $P(X_{iN} = 0) = G_N(P(Y_{i1} = 0))$ are particular cases, correspondingly of (7) and (8), applied for $k = 1$. Equality (10) is a consequence of the Total probability formula. (11) and (12) are Wald's equalities, applied for the coordinates in our model. Formulae for double expectation and covariance entail (13). (14) is a consequence of the formula for correlation, (13) and (12).
4. Definition about conditional probability and (8), entail (15). Replacement of (8) and (9) in formula (34.48), Johnson et. al [7] gives (16). One of the main properties of p.g.fs. is that its first derivative in 1 is equal to the expectation of the r.v. ((34.15), Johnson et. al [7]). Its application, together with (16) gives (17).
5. Formula for correlation, together with (2) and (12) prove (18).
6. Following e.g. p.327 ([6]) we apply (3) for $z_1 = z_2 = \dots = z_k = z$ and obtain (19). The main properties of p.g.f., see e.g. p.327 in [6], together with (19) entails (20).
7. Let $m = 0, 1, \dots, s$.

$$\begin{aligned}
P(X_{iN} = m, X_{1N} + \dots + X_{i-1N} + X_{i+1N} \dots + X_{kN} = j) &= \\
&= \frac{\partial^{m+j} G_{X_{iN}, X_{1N} + \dots + X_{i-1N} + X_{i+1N} \dots + X_{kN}}(z_i, z)}{m!j! \partial z_i^m \partial z^j} \Big|_{(z_i, z) = (0,0)} = \\
&= \frac{\partial^{m+j} G_N(G_{Y_{11}, \dots, Y_{i-11}, Y_{i1}, Y_{i+11}, \dots, Y_{k1}}(z, \dots, z, z_i, z, \dots, z))}{m!j! \partial z_i^m \partial z^j} \Big|_{(z_i, z) = (0,0)}.
\end{aligned}$$

The last equality, together with the definition for conditional probability and (20), entail

$$\begin{aligned}
P(X_{iN} = m | X_{1N} + X_{2N} + \dots + X_{kN} = s) &= \\
&= \frac{P(X_{iN} = m, X_{1N} + \dots + X_{i-1N} + X_{i+1N} \dots + X_{kN} = s - m)}{P(X_{1N} + \dots + X_{i-1N} + X_{i+1N} \dots + X_{kN} = s - m)} = \\
&= \binom{s}{m} \frac{\frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{i-11}, Y_{i1}, Y_{i+11}, \dots, Y_{k1}}(z, \dots, z, z_i, z, \dots, z))}{\partial z^{s-m} \partial z_i^m} \Big|_{(z_i, z) = (0,0)}}{\frac{\partial^s G_N(G_{Y_{11}, \dots, Y_{k1}}(z, \dots, z))}{\partial z^s} \Big|_{z=0}}.
\end{aligned}$$

□

Note: 1. The second formula in (12) shows that, the bigger the FIN , or EY_{i1} , or $FIIY_{i1}$ are, the bigger the FIX_{iN} is.

2. The strength of the linear dependence between X_{iN} and N , expressed in (14), decreases with the increments of FIN or CVY_{i1} .

3. Particular cases

Here we apply the results from Section 2 in some particular cases of compounds with equal number of Mn or NMn distributed summands. We make a brief empirical study of their mean square regressions and observe that in any of these multivariate random sums it is close to linear.

3.1. Multivariate compounds with Multinomial summands.

Consider a partition A_1, A_2, \dots, A_k of the sample space Ω . Mn distribution is the one of the numbers of outcomes A_i , $i = 1, 2, \dots, k$ in a series of n independent repetitions of a trial. More precisely, let $n \in \mathbb{N}$, $0 < p_i$, $i = 1, 2, \dots, k$ and $p_1 + p_2 + \dots + p_k = 1$. A vector $(\xi_1, \xi_2, \dots, \xi_k)$ is called Multinomially distributed with parameters n, p_1, p_2, \dots, p_k , if its probability mass function (p.m.f.) is

$$P(\xi_1 = i_1, \xi_2 = i_2, \dots, \xi_k = i_k) = \binom{n}{i_1, i_2, \dots, i_k} p_1^{i_1} p_2^{i_2} \dots p_k^{i_k},$$

for $i_1 + i_2 + \dots + i_k = n$, $i_s = 0, 1, \dots$, $s = 1, 2, \dots, k$. Briefly

$$(\xi_1, \xi_2, \dots, \xi_k) \sim Mn(n; p_1, p_2, \dots, p_k).$$

The distribution of any subset $\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_r}$ of its coordinates, can be easily described using relation that

$$(22) \quad (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_r}, n - (\xi_{i_1} + \xi_{i_2} + \dots + \xi_{i_r})) \sim Mn(n; p_{i_1}, p_{i_2}, \dots, p_{i_r}, 1 - (p_{i_1} + p_{i_2} + \dots + p_{i_r})).$$

A systematic investigation of this distribution, together with a very good list of references, could be found e.g. in Johnson et al. [7]. Considering this distribution for $k = 1$, we obtain Binomial(Bi) distribution. For $n = 1$ the Mn distribution is multivariate Bernoulli distribution. Some of the following results, for the particular case when N is Poisson distributed, can be found in Smith [14]. This distribution is partially investigated on in 2002, Daley et. al [5], p. 113.

Theorem 2. Suppose $(Y_{1i}, Y_{2i}, \dots, Y_{ki})$, $i = 1, 2, \dots$ are independent Mn distributed random vectors with parameters $(s, p_1, p_2, \dots, p_k)$, where $s \in \mathbb{N}$, $p_1 > 0, p_2 > 0, \dots, p_k > 0$ and $p_1 + p_2 + \dots + p_k = 1$. Denote the resulting distribution, defined in (1) by $(X_{1N}, X_{2N}, \dots, X_{kN}) \sim C N M n(\vec{a}_N; s, p_1, p_2, \dots, p_k)$, where the vector \vec{a}_N contains the parameters of the distribution of the r.v. N . $(X_{1N}, X_{2N}, \dots, X_{kN})$ possesses the following properties.

1. $G_{X_{1N}, X_{2N}, \dots, X_{kN}}(z_1, z_2, \dots, z_k) = G_{sN}(p_1 z_1 + p_2 z_2 + \dots + p_k z_k)$.
2. For $x_1, x_2, \dots, x_k = 0, 1, \dots$, such that $x_1 + x_2 + \dots + x_k$ is a multiple of s ,
 $P(X_{1N} = x_1, X_{2N} = x_2, \dots, X_{kN} = x_k) =$

$$(23) \quad = \binom{x_1 + x_2 + \dots + x_k}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} P(Ns = x_1 + x_2 + \dots + x_k).$$

$$P(X_{1N} = 0, X_{2N} = 0, \dots, X_{kN} = 0) = P(N = 0).$$

3. For all $r = 1, 2, \dots, k$ and for any subset of coordinates $X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}$,
 $G_{X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}}(z_{i_1}, z_{i_2}, \dots, z_{i_r}) =$

$$(24) \quad = G_{sN}(p_{i_1} z_{i_1} + p_{i_2} z_{i_2} + \dots + p_{i_r} z_{i_r} + 1 - (p_{i_1} + \dots + p_{i_r})).$$

$$P(X_{i_1N} = x_{i_1}, X_{i_2N} = x_{i_2}, \dots, X_{i_rN} = x_{i_r}) =$$

$$(25) \quad = \frac{p_{i_1}^{x_{i_1}} \dots p_{i_r}^{x_{i_r}}}{x_{i_1}! x_{i_2}! \dots x_{i_r}!} \sum_{m=0}^{\infty} \frac{(m + x_{i_1} + x_{i_2} + \dots + x_{i_r})!}{m!} \cdot (1 - (p_{i_1} + \dots + p_{i_r}))^m P[(sN = m + x_{i_1} + x_{i_2} + \dots + x_{i_r})],$$

$$(X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}, Ns - (X_{i_1N} + X_{i_2N} + \dots + X_{i_rN})) \sim C N M n(\vec{a}_N; s, p_{i_1}, p_{i_2}, \dots, p_{i_r}, 1 - (p_{i_1} + \dots + p_{i_r})).$$

4. For all $i = 1, 2, \dots, k$, and $m = 0, 1, \dots$,

$$(26) \quad P(X_{iN} = m) = \frac{p_i^m}{m!} \sum_{j=0}^{\infty} \frac{(j+m)!}{j!} (1-p_i)^j P(sN = j+m) =$$

$$(27) \quad = \begin{cases} G_{sN}(1-p_i) & , \quad m = 0 \\ \sum_{n=1}^{\infty} \binom{ns}{m} p_i^m (1-p_i)^{ns-m} P(N=n) & , \quad m = 1, 2, \dots \end{cases}$$

If $EN < \infty$, $EX_{iN} = sp_i EN$.

If $E(N^2) < \infty$, then $Var X_{iN} = sp_i[sp_i Var N + (1 - p_i)EN]$ and

$$(28) \quad FI X_{iN} = 1 + p_i(sFIN - 1).$$

5. $P(X_{iN} = x_i | X_{jN} = x_j) =$

$$(29) \quad = \frac{p_i^{x_i} \sum_{m=0}^{\infty} \frac{(m+x_i+x_j)!}{m!} (1 - (p_i + p_j))^m P(sN = m + x_i + x_j)}{\sum_{m=0}^{\infty} \frac{(x_j+m)!}{m!} (1 - p_j)^m P(sN = x_j + m)},$$

$G_{X_{iN}}(z_i | X_{jN} = x_j) =$

$$(30) \quad = \frac{\sum_{m=0}^{\infty} \frac{(x_j+m)!}{m!} (p_i z_i + 1 - p_i - p_j)^m P(sN = x_j + m)}{\sum_{m=0}^{\infty} \frac{(x_j+m)!}{m!} (1 - p_j)^m P(sN = x_j + m)}.$$

(31)

$$E(X_{iN} | X_{jN} = x_j) = p_i \frac{\sum_{m=0}^{\infty} \frac{(x_j+m+1)!}{m!} (1 - p_j)^m P(sN = x_j + m + 1)}{\sum_{m=0}^{\infty} \frac{(x_j+m)!}{m!} (1 - p_j)^m P(sN = x_j + m)}$$

$$cov(X_{iN}, X_{jN}) = sp_i p_j [-EN + s Var N]$$

$$cor(X_{iN}, X_{jN}) =$$

$$(32) \quad = \frac{\sqrt{p_i p_j} (s Var N - EN)}{\sqrt{[sp_i Var N + EN(1 - p_i)][sp_j Var N + EN(1 - p_j)]}} =$$

$$(33) \quad = \sqrt{\frac{(FI X_{iN} - 1)(FI X_{jN} - 1)}{FI X_{iN} FI X_{jN}}}.$$

6. If the following expectation exists, then

$$(34) \quad E(X_{iN} | X_{jN} = x_j) = \frac{p_i}{p_j} (x_j + 1) \frac{P(X_{jN} = x_j + 1)}{P(X_{jN} = x_j)}.$$

7. $G_{X_{1N} + X_{2N} + \dots + X_{kN}}(z) = G_{sN}(z),$

$$P(X_{1N} + X_{2N} + \dots + X_{kN} = m) = P(sN = m).$$

8. For $j = 1, 2, \dots, i = 1, 2, \dots, k$, $(X_{iN} | X_{1N} + X_{2N} + \dots + X_{kN} = j) \sim Bi(j; p_i)$.

Proof.

1. The form of p.g.f. follows by (3) and the formula of the p.g.f. of Mn distribution. See e.g. (35.4) in [7].
2. This property is an immediate consequence of (4), (5), the fact that

$$X_{1N} + X_{2N} + \dots + X_{kN} = sN,$$

and relation between p.m.f. and p.g.f., see e.g. (34.2) in [7].

3. Here we replace the redundant variables in Equality 1), Th. 2 with 1, use one of the main properties of p.g.f. (see e.g. p.327 in [6] or Daley et. al [5], p. 113) and obtain (24). Using (22) and the definition (1), we have that

$$\begin{aligned} (X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}, Ns - (X_{i_1N} + X_{i_2N} + \dots + X_{i_rN})) &\sim \\ &\sim C N Mn(\vec{a}_N; s, p_{i_1}, p_{i_2}, \dots, p_{i_r}, 1 - (p_{i_1} + \dots + p_{i_r})). \end{aligned}$$

Now we sum up over the possible values of the last coordinate of this vector, use 2), and obtain

$$\begin{aligned} P(X_{i_1N} = x_{i_1}, X_{i_2N} = x_{i_2}, \dots, X_{i_rN} = x_{i_r}) &= \\ = \sum_{m=0}^{\infty} P(X_{i_1N} = x_{i_1}, X_{i_2N} = x_{i_2}, \dots, X_{i_rN} = x_{i_r}, Ns - (X_{i_1N} + X_{i_2N} + \dots & \\ \dots + X_{i_rN}) = m) = & \\ = \sum_{m=0}^{\infty} \binom{x_{i_1} + x_{i_2} + \dots + x_{i_r} + m}{x_{i_1}, x_{i_2}, \dots, x_{i_r}, m} p_{i_1}^{x_{i_1}} \dots p_{i_r}^{x_{i_r}} [1 - (p_{i_1} + \dots + p_{i_r})]^m & \\ \cdot P(Ns = x_{i_1} + x_{i_2} + \dots + x_{i_r} + m) = & \\ = \frac{p_{i_1}^{x_{i_1}} \dots p_{i_r}^{x_{i_r}}}{x_{i_1}! x_{i_2}! \dots x_{i_r}!} \sum_{m=0}^{\infty} \frac{(m + x_{i_1} + x_{i_2} + \dots + x_{i_r})!}{m!} [1 - (p_{i_1} + \dots + p_{i_r})]^m & \\ \cdot P(sN = m + x_{i_1} + x_{i_2} + \dots + x_{i_r}). & \end{aligned}$$

4. The first equality in (26) is a particular case of (25), for $r = 1$. The rest part of (26) is a consequence of the substitution $j + m = ns$ and the definition of p.g.f. The results about the numerical characteristics follow by the Wald's equalities and (22). More precisely $EY_{i1} = sp_i$ and $VarY_{i1} = sp_i(1 - p_i)$.

5. The definitions for conditional probability and (25) imply (29). By (16)

$$\begin{aligned}
G_{X_{iN}}(z_i | X_{jN} = x_j) &= \\
&= \frac{\frac{\partial^{x_j} G_{sN}(p_i z_i + p_j z_j + 1 - p_i - p_j)}{\partial z_j^{x_j}}}{\frac{\partial^{x_j} G_{sN}(p_j z_j + 1 - p_j)}{\partial z_j^{x_j}}} \Big|_{z_j=0} = \frac{p_j^{x_j} \left(\frac{\partial^{x_j}}{\partial y^{x_j}} G_{sN}(y) \right) \Big|_{y=p_i z_i + 1 - p_i - p_j}}{p_j^{x_j} \left(\frac{\partial^{x_j}}{\partial y^{x_j}} G_{sN}(y) \right) \Big|_{y=1-p_j}} = \\
&= \frac{\sum_{n=0}^{\infty} \left(\frac{\partial^{x_j}}{\partial y^{x_j}} y^{sn} P(N = n) \right) \Big|_{y=p_i z_i + 1 - p_i - p_j}}{\sum_{n=0}^{\infty} \left(\frac{\partial^{x_j}}{\partial y^{x_j}} y^{sn} P(N = n) \right) \Big|_{y=1-p_j}} = \\
&= \frac{\sum_{n: sn \geq x_j} \frac{(sn)!}{(sn-x_j)!} (p_i z_i + 1 - p_i - p_j)^{sn-x_j} P(sN = sn)}{\sum_{n: sn \geq x_j} \frac{(sn)!}{(sn-x_j)!} (1-p_j)^{sn-x_j} P(sN = sn)}.
\end{aligned}$$

After the Substitution $m = sn - x_j$ we obtain (30).

In order to calculate covariation we use (2) and the formula for covariance of Mn random vector $cov(Y_{i1}, Y_{j1}) = -np_i p_j$. See e.g. Johnson et. al [7]. Together with $Var X_{iN} = sp_i [sp_i Var N + (1 - p_i) EN]$ and formula for the correlation they imply (32). Divide the numerator and denominator in $cor(X_{iN}, X_{jN})$ to EN and come to

$$\frac{\sqrt{p_i p_j} (sFIN - 1)}{\sqrt{[sp_i FIN + 1 - p_i][sp_j FIN + 1 - p_j]}}.$$

Compare this result with (28) we see that it is equal to (33).

6. By (31) and (26),

$$\begin{aligned}
E(X_{iN} | X_{jN} = x_j) &= p_i \frac{\sum_{m=0}^{\infty} \frac{(x_j+m+1)!}{m!} (1-p_j)^m P(sN = x_j + m + 1)}{\sum_{m=0}^{\infty} \frac{(x_j+m)!}{m!} (1-p_j)^m P(sN = x_j + m)} = \\
&= p_i \frac{\frac{(x_j+1)!}{p_j^{x_j+1}} P(X_{jN} = x_j + 1)}{\frac{x_j!}{p_j^{x_j}} P(X_{jN} = x_j)} = \frac{p_i}{p_j} (x_j + 1) \frac{P(X_{jN} = x_j + 1)}{P(X_{jN} = x_j)}.
\end{aligned}$$

7. These distributions are obtained by 1) Th. 2, (19), (20) and the properties of p.g.fs.

8. By (1), we have $(X_{iN}, X_{1N} + X_{2N} + \dots + X_{i-1N} + X_{i+1N} + \dots + X_{kN}) \sim \sim C N M n(\vec{a}_N; s, p_i, 1 - p_i)$. This, together with the definitions for conditional probability, (23) and Mn (via its p.m.f.) entail

$$\begin{aligned}
 P(X_{iN} = m | X_{1N} + X_{2N} + \dots + X_{kN} = j) &= \\
 &= \frac{P(X_{iN} = m, X_{1N} + \dots + X_{i-1N} + X_{i+1N} + \dots + X_{kN} = j - m)}{P(X_{1N} + X_{2N} + \dots + X_{kN} = j)} = \\
 &= \frac{\binom{j}{m} p_i^m (1 - p_i)^{j-m} P(Ns = j)}{P(Ns = j)} = \binom{j}{m} p_i^m (1 - p_i)^{j-m}, \\
 &\qquad\qquad\qquad m = 0, 1, \dots, j.
 \end{aligned}$$

Now, the definition of Bi distribution entails the desired result.

□

Note: 1. As corollaries of the above theorem, when we fix the distribution of N , we can obtain many new and well known distributions and their numerical characteristics and conditional distributions.

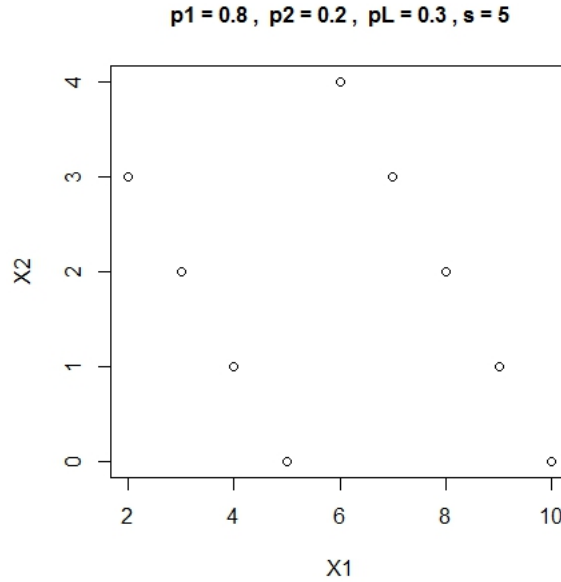


Fig. 1.

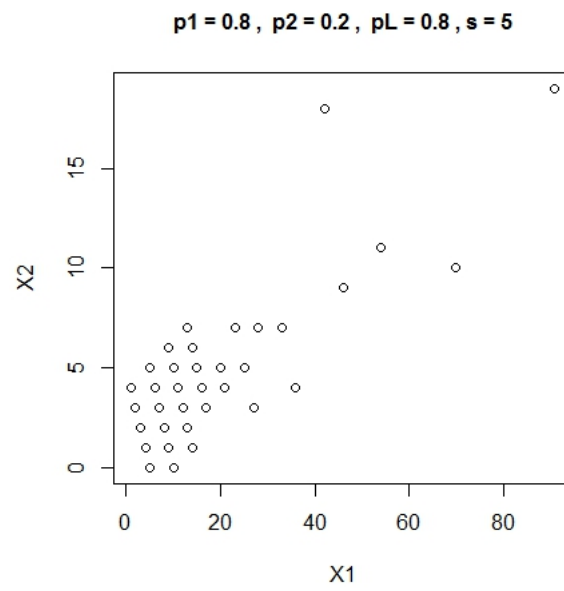


Fig. 2.

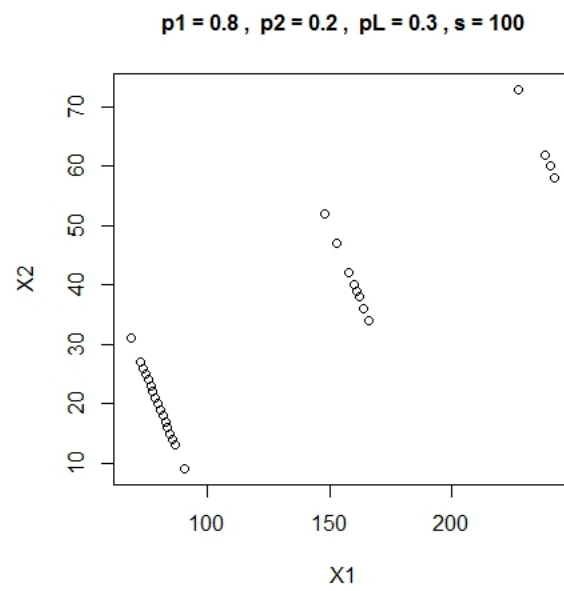


Fig. 3.

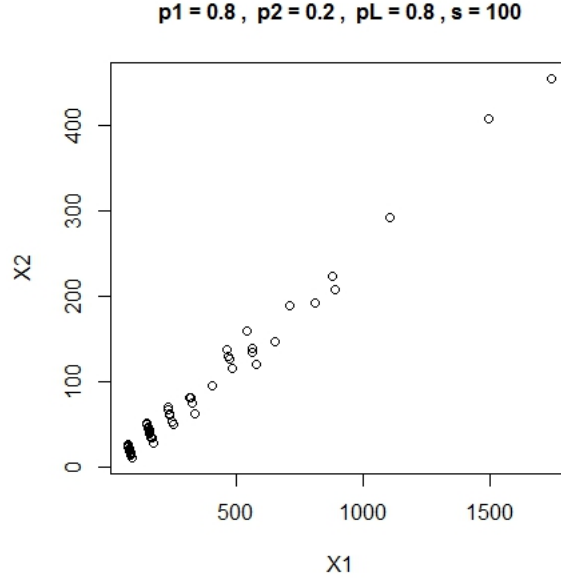


Fig. 4.

2. Here we denote by $LS(p_L)$ the Logarithmic series distribution with parameter $p_L \in (0, 1)$. In order to observe the form of the dependence between the r.v.s. X_{1N} , X_{2N} and to model it via their mean square regression (34), we simulate 100 realizations of $(X_{1N}, X_{2N}) \sim CLSMn(p_L; s, p_1, p_2)$, for different values of the parameters p_L , s , p_1 and p_2 . Four of the scatter plots of the data are given on Figures 1 - 4. These observations show that the dependence between these two r.v.s. is approximately linear, with slope almost $\frac{p_i}{p_j}$. The relation (33) entails that the strength of the dependence between X_{iN} and X_{jN} is governed by $FI X_{iN}$ and $FI X_{jN}$. When the last two Fisher indexes increase, the correlation between X_{iN} and X_{jN} also increases. Equation (28) entails that the bigger the FIN , or s are, the bigger the $FI X_{iN}$ is. The last increases the strength of the dependence between X_{iN} and X_{jN} . All these results were observed empirically. More precisely Figures 1 - 4, show that for bigger s or p_L we obtain stronger linear dependence between X_{iN} and X_{jN} .

3. The comparison of the coefficients before $P(N = n)$ in both expressions for the p.m.f. of X_{iN} , $i = 1, 2, \dots, k$, (the first obtained using the Total probability

formula and the second is (27)) gives us another proof of the equality

$$\sum_{j_1+j_2+\dots+j_n=m} \binom{s}{j_1} \binom{s}{j_2} \dots \binom{s}{j_n} = \binom{ns}{m}.$$

In analogous way, working with the theoretical results, related with multivariate compounds with equal number of summands we can obtain new similar equalities.

3.2. Compounds with NMn summands

In this subsection we consider the case, when the summands in (1) are NMn. It seems that in 1952, Bates et. al [1] introduced NMn distribution. They reached it considering Poisson distribution with Gamma mixing variable. Their first parameter could be a real number. Wishart [19] investigated the case, when the first parameter could be only integer. He called this distribution Pascal multinomial distribution. From practical point of view, the NMn distribution is interpreted as the one of the numbers of outcomes A_i , $i = 1, 2, \dots, k$ before the n -th B , in series of independent repetitions, where A_i , $i = 1, 2, \dots, k$ and B form a partition of the sample space. Let us remind the definition.

Let $n \in \mathbb{N}$, $0 < p_i$, $i = 1, 2, \dots, k$ and $p_1 + p_2 + \dots + p_k < 1$. A vector $(\xi_1, \xi_2, \dots, \xi_k)$ is called Negative multinomially distributed with parameters n, p_1, p_2, \dots, p_k , if its probability mass function (p.m.f.) is

$$\begin{aligned} P(\xi_1 = i_1, \xi_2 = i_2, \dots, \xi_k = i_k) = \\ = \binom{n + i_1 + i_2 + \dots + i_k - 1}{i_1, i_2, \dots, i_k, n - 1} p_1^{i_1} p_2^{i_2} \dots p_k^{i_k} (1 - p_1 - p_2 - \dots - p_k)^n, \end{aligned}$$

$i_s = 0, 1, \dots, s = 1, 2, \dots, k$. Briefly $(\xi_1, \xi_2, \dots, \xi_k) \sim NMn(n; p_1, p_2, \dots, p_k)$.

We will use the distribution of its coordinates. For $m = 2, 3, \dots, k - 1$,

$$(35) \quad (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_r}) \sim NMn(s; \rho_{i_1}, \rho_{i_2}, \dots, \rho_{i_r}),$$

with $\rho_{i_m} = \frac{p_{i_m}}{1 - \sum_{j \notin \{i_1, i_2, \dots, i_r\}} p_j}$, $m = 1, 2, \dots, r$. More properties of NMn distribution can be found in Johnson et al. [7].

If A_1, A_2, \dots, A_k describe all possible mutually exclusive "successes" and the event $\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_k}$ presents the "failure", then the coordinates ξ_i of the above vector, can be interpreted as the number of "successes" of type A_i $i = 1, 2, \dots, k$ until n -th "failure". Considering this distribution for $k = 1$, we obtain NBi distribution with parameters s and $(1 - p_1)$. Let us note that if $k = 2, 3, \dots$

the marginal distributions of NMn distributed random vector, defined above are $NBi(n, 1 - \rho_i)$, $\rho_i = \frac{p_i}{1 - \sum_{j \neq i} p_j}$. More precisely their p.g.f. is

$$G_\xi(z) = Ez^{\xi_i} = \left(\frac{1 - \rho_i}{1 - \rho_i z} \right)^n, \quad |z| < \frac{1}{\rho_i} \quad i = 1, 2, \dots, k.$$

Having in mind the univariate distributions, the NMn distribution is sometimes called Multivariate Negative Binomial distribution. For $n = 1$ the NMn distribution is Multivariate geometric distribution. Some properties of the bivariate version of this distribution are considered e.g. in 1981, by Phatak et al. [13]. A systematic investigation of multivariate version could be found e.g. in Srivastava et al. [15]. Here we consider random sums with Mn distributed summands. The case when N is Poisson distributed is considered in Smith [14]. The next theorem generalizes his results.

Theorem 3. Suppose $(Y_{1i}, Y_{2i}, \dots, Y_{ki}) \sim NMn(s, p_1, p_2, \dots, p_k)$, $i = 1, 2, \dots$ are independent. Here $s \in \mathbb{N}$, $p_1 > 0, p_2 > 0, \dots, p_k > 0$ and

$$p_1 + p_2 + \dots + p_k < 1.$$

Denote by $p_0 = 1 - p_1 - p_2 - \dots - p_k$. Then the resulting distribution of

$$(X_{1N}, X_{2N}, \dots, X_{kN}) \sim C N NMn(\vec{a}_N; s, p_1, p_2, \dots, p_k),$$

defined in (1), possesses the following properties.

1. Its p.g.f. is

(36)

$$G_{X_{1N}, X_{2N}, \dots, X_{kN}}(z_1, z_2, \dots, z_k) = G_{sN} \left(\frac{p_0}{1 - p_1 z_1 - p_2 z_2 - \dots - p_k z_k} \right).$$

2. The p.m.f. satisfies the equalities:

$$P(X_{1N} = 0, X_{2N} = 0, \dots, X_{kN} = 0) = G_{sN}(p_0).$$

For $s = 1, 2, \dots, k$, $x_s = 0, 1, \dots$: $(x_1, x_2, \dots, x_k) \neq (0, 0, \dots, 0)$,

$$P(X_{1N} = x_1, X_{2N} = x_2, \dots, X_{kN} = x_k) =$$

$$(37) \quad = p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} \sum_{j=1}^{\infty} \binom{s_j + x_1 + x_2 + \dots + x_k - 1}{x_1, x_2, \dots, x_k, s_j - 1} p_0^{js} P(N = j).$$

3. For all $r = 1, 2, \dots, k$ and for all $X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}$,

$$\begin{aligned} G_{X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}}(z_{i_1}, z_{i_2}, \dots, z_{i_r}) &= \\ (38) \quad &= G_{sN} \left(\frac{\rho_{0r}}{1 - \rho_{i_1} z_{i_1} - \rho_{i_2} z_{i_2} - \dots - \rho_{i_r} z_{i_r}} \right), \end{aligned}$$

where $\rho_{i_j} = \frac{p_{i_j}}{p_0 + \sum_{j=1}^r p_{i_j}}$, $\rho_{0r} = 1 - \rho_{i_1} - \dots - \rho_{i_r}$, i.e.

$$(X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}) \sim C N N M n(\vec{a}_N; s, \rho_{i_1}, \rho_{i_2}, \dots, \rho_{i_r}).$$

$$P(X_{i_1N} = 0, X_{i_2N} = 0, \dots, X_{i_rN} = 0) = G_{sN}(\rho_{0r}).$$

For $s = 1, \dots, r$, $x_{i_s} = 0, 1, \dots$: $(x_{i_1}, x_{i_2}, \dots, x_{i_r}) \neq (0, 0, \dots, 0)$,

$$\begin{aligned} P(X_{i_1N} = x_{i_1}, X_{i_2N} = x_{i_2}, \dots, X_{i_rN} = x_{i_r}) &= \\ (39) \quad &= \rho_{i_1}^{x_{i_1}} \dots \rho_{i_r}^{x_{i_r}} \sum_{j=1}^{\infty} \binom{s j + x_{i_1} + x_{i_2} + \dots + x_{i_r} - 1}{x_{i_1}, x_{i_2}, \dots, x_{i_r} + s j - 1} \rho_{0r}^{js} P(N = j). \end{aligned}$$

4. If $E N < \infty$, then $E X_{iN} = s \frac{p_i}{p_0} E N$.

If $E(N^2) < \infty$, then $\text{Var } X_{iN} = s \frac{p_i^2}{p_0^2} [s \text{Var } N + (1 + \frac{p_0}{p_i}) E N]$,

$$(40) \quad FI X_{iN} = 1 + \frac{p_i}{p_0} [s FI N + 1].$$

5. Dependence between the coordinates.

$$(41) \quad \text{cor}(X_{iN}, X_{jN}) = \sqrt{\frac{(FI X_{iN} - 1)(FI X_{jN} - 1)}{FI X_{iN} FI X_{jN}}}.$$

6. Its conditional distributions satisfy the next equalities.

For $x_j \neq 0$, $x_i = 0, 1, \dots$,

$$(42) \quad P(X_{iN} = x_i | X_{jN} = x_j) = \frac{1}{x_i!} \left(\frac{p_0 + p_j}{p_0 + p_i + p_j} \right)^{x_j} \left(\frac{p_i}{p_0 + p_i + p_j} \right)^{x_i}.$$

$$\frac{\sum_{n=1}^{\infty} \frac{(sn+x_i+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_i+p_j}\right)^{sn} P(N=n)}{\sum_{n=1}^{\infty} \frac{(sn+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_j}\right)^{sn} P(N=n)}.$$

For $x_i = 1, 2, \dots$, $P(X_{iN} = x_i | X_{jN} = 0) =$

$$= \frac{\left(\frac{p_i}{p_0+p_i+p_j}\right)^{x_i}}{G_{sN}\left(\frac{p_0}{p_0+p_j}\right)} \sum_{n=1}^{\infty} \frac{(sn+x_i-1)!}{x_i!(sn-1)!} \left(\frac{p_0}{p_0+p_i+p_j}\right)^{sn} P(N=n),$$

$$P(X_{iN} = 0 | X_{jN} = 0) = \frac{G_{sN}\left(\frac{p_0}{p_0+p_i+p_j}\right)}{G_{sN}\left(\frac{p_0}{p_0+p_j}\right)}.$$

7. For $x_j = 1, 2, \dots$

$$(43) \quad E(X_{iN} | X_{jN} = x_j) = \frac{p_i}{p_j} (x_j + 1) \frac{P(X_{jN} = x_j + 1)}{P(X_{jN} = x_j)}.$$

8. $X_{1N} + X_{2N} + \dots + X_{kN} \sim CNMn(s; 1 - p_0)$.

9. For $j = 1, 2, \dots$, $i = 1, 2, \dots, k$,

$$(X_{iN} | X_{1N} + X_{2N} + \dots + X_{kN} = j) \sim Bi(j; \frac{p_i}{1 - p_0}).$$

Proof.

1. The equality (36) follows by (3) and the formula of the p.g.f. of NMn distribution. See e.g. (36.1) in [7], who use different parametrisation.
2. The formula of the p.m.f. at $(0, 0, \dots, 0)$ is an immediate consequence of (4), the fact that $G_N(z^s) = G_{sN}(z)$ and the form of p.m.f. of $(Y_{11}, Y_{21}, \dots, Y_{k1}) \sim NMn(s, p_1, p_2, \dots, p_k)$ at $(0, 0, \dots, 0)$.

To prove (37) we can use the Total probability formula (the partition of the sample space describe the possible values of N), and the fact that n -th fold convolution on $NMn(s, p_1, p_2, \dots, p_k)$ is $NMn(ns, p_1, p_2, \dots, p_k)$. The same formula can be obtained also by (5) and relation between p.m.f. and p.g.f.

3. Equality 1), Th. 3 entails (38), when replace redundant variables with 1 and use one of the main properties of p.g.f., see e.g. p.327 in [6]. Using (35) and the definition (1), we have that

$$(X_{i_1N}, X_{i_2N}, \dots, X_{i_rN}) \sim C N N M n(\vec{a}_N; s, \rho_{i_1}, \rho_{i_2}, \dots, \rho_{i_r}).$$

Now we use (39) and complete the proof.

4. The substitution of the mean $EY_{i1} = s \frac{\rho_i}{1-\rho_i} = s \frac{p_i}{p_0}$, and the variance $Var Y_{i1} = s \frac{p_i(p_0+p_i)}{p_0^2}$, $i = 1, 2, \dots, k$, of NMn distribution in (11) and (12) gives us $E X_{iN}$ and $Var X_{iN}$, $i = 1, 2, \dots, k$.
5. Using the properties of NMn distribution we have that

$$[CV(Y_{i1}, Y_{i2})]^2 = \frac{1}{s}, (CV Y_{i1})^2 = \frac{1}{s}(1 + \frac{p_0}{p_i}), (CV Y_{j1})^2 = s^{-1}(1 + \frac{p_0}{p_j}).$$

Now the equation (18), together with (40) entail the desired result.

6. The definitions for conditional probability and (39), applied for $r = 1$ and $r = 2$, imply (42). In analogous way we prove the case when $X_{jN} = 0$.
7. Let x_j be a fixed positive integer. In the next calculations we first use the formula for the mean, then we change the order of summation, use the derivatives of the geometric series, make some algebra and obtain.

$$\begin{aligned} E(X_{iN} | X_{jN} = x_j) &= \sum_{x_j=1}^{\infty} \frac{1}{(x_i - 1)!} \left(\frac{p_0 + p_j}{p_0 + p_i + p_j} \right)^{x_j} \left(\frac{p_i}{p_0 + p_i + p_j} \right)^{x_i} \\ &\quad \cdot \frac{\sum_{n=1}^{\infty} \frac{(sn+x_i+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_i+p_j} \right)^{sn} P(N=n)}{\sum_{n=1}^{\infty} \frac{(sn+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_j} \right)^{sn} P(N=n)} = \\ &= \left(\frac{p_i}{p_0 + p_i + p_j} \right) \sum_{n=1}^{\infty} \frac{P(N=n)}{(sn-1)!} \left(\frac{p_0}{p_0 + p_i + p_j} \right)^{sn} \left(\frac{p_0 + p_j}{p_0 + p_i + p_j} \right)^{x_j} \\ &\quad \cdot \frac{\sum_{x_j=1}^{\infty} \frac{(sn+x_i+x_j-1)!}{(x_i-1)!} \left(\frac{p_i}{p_0+p_i+p_j} \right)^{x_i-1}}{\sum_{n=1}^{\infty} \frac{(sn+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_j} \right)^{sn} P(N=n)} = \\ &= \left(\frac{p_i}{p_0 + p_i + p_j} \right) \sum_{n=1}^{\infty} \frac{P(N=n)}{(sn-1)!} \left(\frac{p_0}{p_0 + p_i + p_j} \right)^{sn} \left(\frac{p_0 + p_j}{p_0 + p_i + p_j} \right)^{x_j} \end{aligned}$$

$$\begin{aligned}
& \frac{\frac{(sn+x_j)!}{\left(1-\frac{p_i}{p_0+p_i+p_j}\right)^{sn+x_j+1}}}{\sum_{n=1}^{\infty} \frac{(sn+x_j-1)!}{(sn-1)!} \left(\frac{p_0}{p_0+p_j}\right)^{sn} P(N=n)} = \\
& = (x_j+1) \frac{p_i}{p_j} \frac{\left(\frac{p_j}{p_0+p_j}\right)^{x_j+1} \sum_{n=1}^{\infty} \frac{P(N=n)(sn+x_j)!}{(sn-1)!(x_j+1)!} \left(\frac{p_0}{p_0+p_j}\right)^{sn}}{\left(\frac{p_j}{p_0+p_j}\right)^{x_j} \sum_{n=1}^{\infty} \frac{(sn+x_j-1)!}{(sn-1)!x_j!} \left(\frac{p_0}{p_0+p_j}\right)^{sn} P(N=n)} = \\
& = \frac{p_i}{p_j} (x_j+1) \frac{P(X_{jN} = x_j+1)}{P(X_{jN} = x_j)}.
\end{aligned}$$

Finally we have just compared the previous expression with $P(X_{jN} = x_j + 1)$ and $P(X_{jN} = x_j)$, using (39), for $r = 1$.

8. These distributions are obtained by (19), 1) Th. 3 for $z_1 = z_2 = \dots = z_k = z$ and (38), for $r = 1$.
9. By (1) and the definition of the NMn distribution, we have

$$\begin{aligned}
(X_{iN}, X_{1N} + X_{2N} + \dots + X_{i-1N} + X_{i+1N} + \dots + X_{kN}) & \sim \\
& \sim C N M n(\vec{a}_N; s, p_i, 1 - p_0 - p_i).
\end{aligned}$$

This, together with the definitions for conditional probability, (23) and Mn (via its p.m.f.) entail

$$\begin{aligned}
P(X_{iN} = m | X_{1N} + X_{2N} + \dots + X_{kN} = j) & = \\
& = \frac{P(X_{iN} = m, X_{1N} + \dots + X_{i-1N} + X_{i+1N} + \dots + X_{kN} = j - m)}{P(X_{1N} + X_{2N} + \dots + X_{kN} = j)} = \\
& = \frac{p_i^m (1 - p_0 - p_i)^{j-m} \sum_{n=1}^{\infty} \binom{sn+j-1}{m, j-m, sn-1} p_0^{ns} P(N=n)}{(1 - p_0)^j \sum_{n=1}^{\infty} \binom{sn+j-1}{j} p_0^{ns} P(N=n)} = \\
& = \binom{j}{m} \left(\frac{p_i}{1 - p_0}\right)^m \left(1 - \frac{p_i}{1 - p_0}\right)^j, \quad m = 0, 1, \dots, j.
\end{aligned}$$

Now we use the definition of Bi distribution and complete the proof.

□

Note: 1. In order to observe the form of the dependence between the r.v.s. X_{1N}, X_{2N} , we have simulated 100 realizations of $(X_{1N}, X_{2N}) \sim C L S N M n(p_L;$

s, p_1, p_2) and have plotted the scatter plots of the observations. Here $LS(p_L)$ means the Logarithmic series distribution with parameter $p_L \in (0, 1)$. Four examples of such scatter plots, for different values of the parameters p_L, s, p_1 and p_2 , are given on Figures 5 - 8. They show that the dependence between these two r.v.s. is approximately linear, with slope, close to $\frac{p_i}{p_j}$. The strength of the dependence increases with parameters s and p_L . These empirical observations are synchronised with the theoretical relations, expressed in the equations (43), (41) and (40).

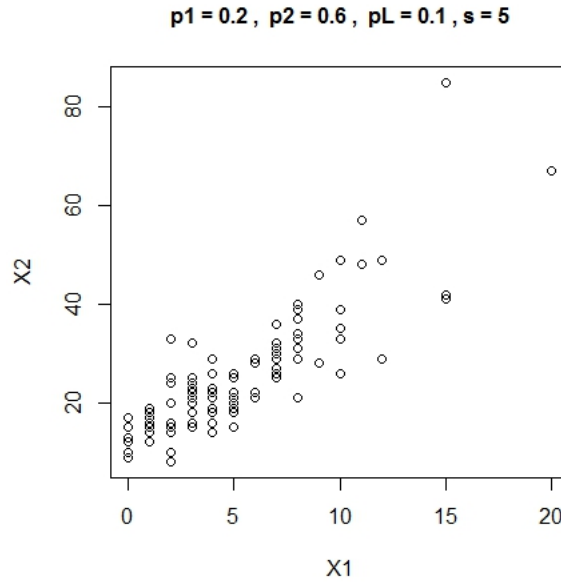


Fig. 5.

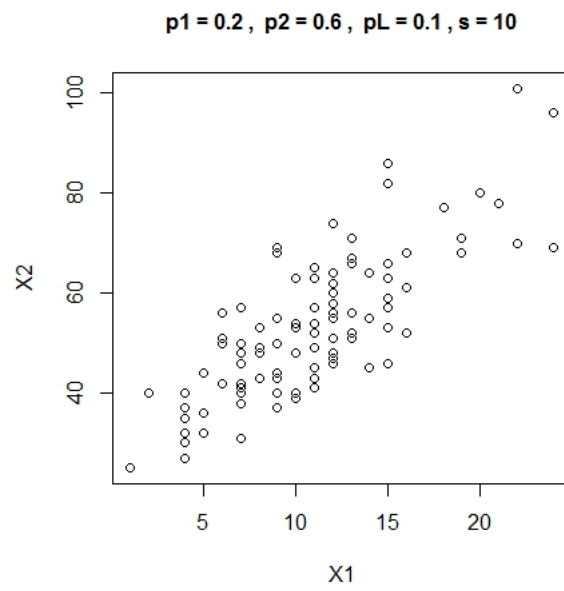


Fig. 6.

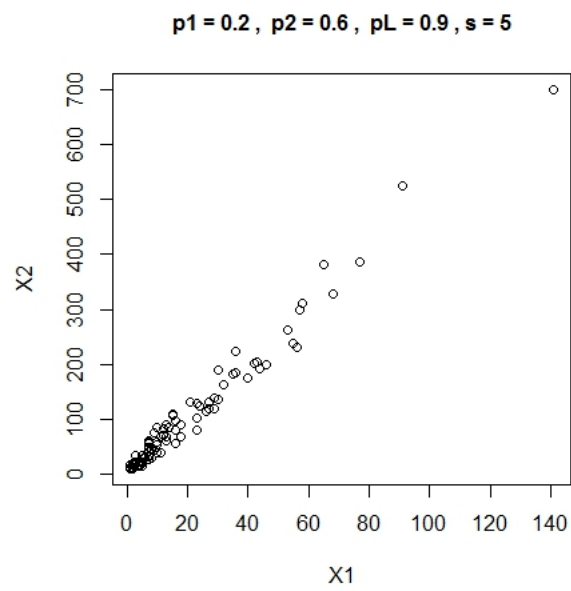


Fig. 7.

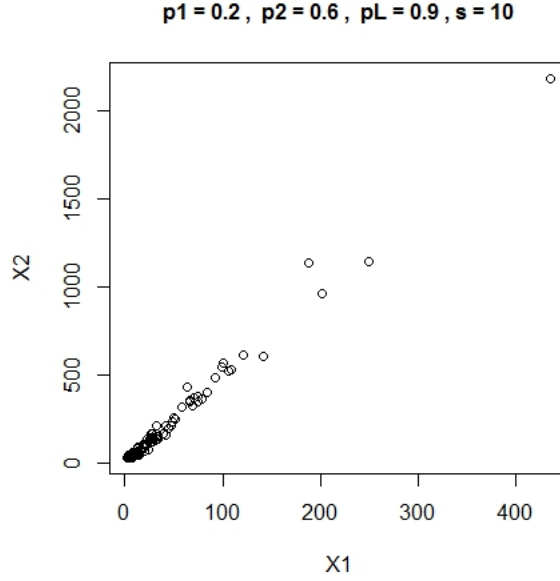


Fig. 8.

2. Similarly to the previous case, the relation (41) shows that the strength of the linear dependence between X_{iN} and X_{jN} is completely governed by the values of $FI X_{iN}$ and $FI X_{jN}$. When the last two Fisher indexes increase, the correlation between X_{iN} and X_{jN} also increases. From the other side, the equation (40) entails that the bigger the $FI N$, or s , or p_0 , or p_i are, the bigger the $FI X_{iN}$ is. So, via $FI X_{iN}$, the last increases the strength of the straight-line dependence between X_{iN} and X_{jN} . On the examples on Figure 3 and Figure 4 we observe that the bigger the s , the stronger the linear dependence is.

4. Conclusive remarks

The paper considers the properties of the distributions of multivariate random sums with equal number of summands. Some simulations of realizations of particular cases of random vectors with such distributions are made, and the dependence between their coordinates is observed. These investigations show that these distributions are appropriate for modelling of both: linear dependence between the coordinates and clustering in the observations. From theoretical point of view, in both particular cases, described here, these results can be seen in equalities (34) and (43). For the general case, without fixing neither the distribution of the number of summands, nor the distribution of summands, the second

formula in (12) shows that, the bigger the Fisher index(FI) of the number of summands, or the expectation, or the FI of the summands are, the bigger the FI of the corresponding random sum is. The strength of the linear dependence between the coordinates X_{iN} of the random sums and the number of summands N , expressed in (14), decreases with the increments of the FI of N or the coefficient of variation of the summands.

If $(X_{1N}, X_{2N}) \sim C L S M n(p_L; s, p_1, p_2)$ the dependence between these two r.vs. is approximately linear, with slope almost $\frac{p_i}{p_j}$. The relation (33) entails that the strength of the dependence between X_{iN} and X_{jN} is governed by $FI X_{iN}$ and $FI X_{jN}$. When the last two Fisher indexes increase, the correlation between X_{iN} and X_{jN} also increases. Equation (28) entails that the bigger the $FI N$, or s are, the bigger the $FI X_{iN}$ is. Figures 1 - 4 show that for bigger parameters s or p_L we obtain stronger linear dependence between X_{iN} and X_{jN} .

It is interesting that in the other, considered particular case here, i.e. if $(X_{1N}, X_{2N}) \sim C L S N M n(p_L; s, p_1, p_2)$, the form of the dependence between X_{1N} and X_{2N} is again approximately linear, with slope almost $\frac{p_i}{p_j}$, for all possible discrete distributions of N . This is synchronised with (43). The strength of the dependence increases with s and p_L . The equation (40) entails that the bigger the $FI N$, or s , or p_0 , or p_i are, the bigger the $FI X_{iN}$ is. Similarly to the previous case, the relation (41) shows that the strength of the linear dependence between X_{iN} and X_{jN} is completely governed by the values of $FI X_{iN}$ and $FI X_{jN}$. When the last two Fisher indexes increase, the correlation between X_{iN} and X_{jN} also increases.

REFERENCES

- [1] BATES, G. E., NEYMAN, J., (1952) Contributions to the Theory of Accident Proneness. 1. An Optimistic Model of the Correlation Between Light and Severe Accidents, *California Univ. Berkeley* **132**.
- [2] CHARALAMBIDES, C. A. (1981) Bivariate Generalized discrete distributions and bipartitional polynomials, *Statistical distributions in scientific work, Springer Netherlands* **4** 213 – 223.
- [3] CHARALAMBIDES, CH. A., PAPAGEORGIOU, H. (1981) Bivariate Poisson binomial distributions, *Biometrical Journal* **23(5)** 437–450.
- [4] COLIN R. B. (1972) On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*. **67 (338)** 364-366.

- [5] DALEY, D. J. , VERE-JONES D. (2002) An Introduction to the Theory of Point Processes: Volume 1: Elementary Theory and Methods. *Springer New York, Berlin, Heidelberg, Hong Kong, London, Milan, Paris, Tokyo*
- [6] DENUIT, M., MARECHAL, X., PITREBOIS, S., WALHIN, J. F. (2007). Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems. *John Wiley and Sons*.
- [7] JOHNSON, N.L., KOTZ, S., BALAKRISHNAN, N. (1997) Discrete multivariate distributions, *Wiley, New York* **165**
- [8] JORDANOVA, P. K. (2016) On multivariate discrete compound distributions with one and the same number of summands, *Proceedings of 2016, Shumen, Bulgaria* **1** 127 – 135.
- [9] KHATRI, C. G. (1971) On multivariate contagious distributions. *Sankhya: The Indian Journal of Statistics Series B* **197–216**.
- [10] LUNDBERG, O. (1940) On random processes with their application to sickness and accident statistics, *Doctoral thesis monograph*.
- [11] MINKOVA, L. D., BALAKRISHNAN, N. (2014) Type II bivariate Polya-Aeppli distribution, *Statistics and Probability Letters* **88** 40–49.
- [12] MINKOVA, L. D., BALAKRISHNAN, N. (2014) On a bivariate Polya-Aeppli distribution, *Communications in Statistics-Theory and Methods* **43.23** 5026–5038.
- [13] PHATAK, A. G., SREEHARI, M. (1981) Some characterizations of a bivariate geometric distribution. *Journal of Indian Statistical Association, Pune* **19** 141–146.
- [14] SMITH, G. E. J. (1965) Parameter estimation in some multivariate compound distributions, *PhD Thesis, The University of British Columbia, Canada*.
- [15] SRIVASTAVA, R. C., BAGCHI, S. N. (1985) On some characterizations of the univariate and multivariate geometric distributions. *Journal of Indian Statistical Association, Pune* **23** 27–33.
- [16] SUNDT, B., VERNIC, R., (2009) Multivariate Compound Distributions of Type 1. *Recursions for Convolutions and Compound Distributions with Insurance Applications. Springer Berlin Heidelberg*, 245–258.

- [17] WALD, A. (1944) On cumulative sums of random variables, *The Annals of Mathematical Statistics* **15(3)** 283–296.
- [18] WANG, Y. H. (1996) On the multivariate compound distributions, *Journal of multivariate analysis* **59(1)** 13–21.
- [19] WISHART, J. (1949) Cumulants of multivariate multinomial distribution, *Biometrika, Oxford University Press* **36** 47–58.

Pavlina Kalcheva Jordanova
Shumen University
115 Universitetska str.,
9700 Shumen, Bulgaria
e-mail: pavlina.kj@abv.bg

